Research Article | Vol. 12, No. 12 / 1 Dec 2021 / Biomedical Optics Express 7348
Biomedical Optics EXPRESS

# Multi-task learning approach for volumetric segmentation and reconstruction in 3D OCT images

DHEO A. Y. CAHYO,[1,2] AI PING YOW,[1,2,3] SEANG-MEI SAW,[2] MARCUS ANG,[2] MICHAEL GIRARD,[2] LEOPOLD SCHMETTERER,[1,2,4,5,6,7] AND DAMON WONG[1,2,4,7,*]

[1] *SERI-NTU Advanced Ocular Engineering (STANCE), Singapore*
[2] *Singapore Eye Research Institute, Singapore National Eye Centre, Singapore*
[3] *Singapore Centre for Environmental Life Sciences Engineering (SCELSE), Singapore*
[4] *School of Chemical and Biomedical Engineering, Nanyang Technological University, Singapore*
[5] *Department of Clinical Pharmacology, Medical University of Vienna, Austria*
[6] *Center for Medical Physics and Biomedical Engineering, Medical University of Vienna, Austria*
[7] *Institute of Molecular and Clinical Ophthalmology, Basel, Switzerland*
[*] *damon.wong@ntu.edu.sg*

**Abstract:** The choroid is the vascular layer of the eye that supplies photoreceptors with oxygen. Changes in the choroid are associated with many pathologies including myopia where the choroid progressively thins due to axial elongation. To quantize these changes, there is a need to automatically and accurately segment the choroidal layer from optical coherence tomography (OCT) images. In this paper, we propose a multi-task learning approach to segment the choroid from three-dimensional OCT images. Our proposed architecture aggregates the spatial context from adjacent cross-sectional slices to reconstruct the central slice. Spatial context learned by this reconstruction mechanism is then fused with a U-Net based architecture for segmentation. The proposed approach was evaluated on volumetric OCT scans of 166 myopic eyes acquired with a commercial OCT system, and achieved a cross-validation Intersection over Union (IoU) score of 94.69% which significantly outperformed ($p<0.001$) the other state-of-the-art methods on the same data set. Choroidal thickness maps generated by our approach also achieved a better structural similarity index (SSIM) of 72.11% with respect to the groundtruth. In particular, our approach performs well for highly challenging eyes with thinner choroids. Compared to other methods, our proposed approach also requires lesser processing time and has lower computational requirements. The results suggest that our proposed approach could potentially be used as a fast and reliable method for automated choroidal segmentation.



## 1. Introduction

The choroid is the vascular layer between the sclera and the retina which plays an important role in transporting oxygen and nutrients to the eye outer retina including the photoreceptors. [1] Many eye diseases such as myopia and age-related macular degeneration (AMD), have been associated with choroidal changes. [2–8] Studies have been conducted through the use of non-invasive imaging techniques to acquire images of the subject's retina so as to investigate the choroidal changes in these images. [2] For example, choroidal thinning is associated with myopia and axial elongation. [9] Many studies in myopia showed this correlation with sub-foveal choroidal thickness changes. [3–5] Flores-Moreno et al. [5] reported that in eyes with high myopia, a millimeter increase in the axial length was associated with a choroid thickness decrease of $25.9\boldsymbol{\mu m} \pm 2.1\boldsymbol{\mu m}$. Other examples of eye conditions associated with choroidal layer changes are AMD [6], where the choroid thickness has been shown to be correlated with severity of

#428140 https://doi.org/10.1364/BOE.428140
Journal © 2021 Received 22 Apr 2021; revised 11 Jul 2021; accepted 27 Aug 2021; published 8 Nov 2021

Biomedical Optics EXPRESS

non-exudative macular changes [7], and Diabetic Retinopathy (DR) where increasing thickness of the choroid was shown to be correlated with severity of retinopathy [8]. These studies have demonstrated the importance of detecting and monitoring changes in the choroid.

Optical Coherence Tomography (OCT) is an interferometric technique which uses low-coherence light to allow high-resolution, cross-sectional tomographic imaging of tissue. [10] This non-invasive optical imaging technique has been widely adopted in ophthalmic practice and research to allow visualization and quantification of the structures in the eye. A challenge in OCT imaging is the limited penetration depth due to multiple scattering effects in retinal tissue. Recent swept-source OCT systems have used longer wavelengths, which are less susceptible to scattering and allow deeper penetration depths. However, the visibility of the outer boundary of the choroid remains relatively poorer compared to the other retinal layers.

Segmentation of the choroid in OCT is important to enable detailed analysis of the choroidal layers. [11] However, as manual segmentation of the choroid is labour-intensive and time consuming, there is an interest in developing automated segmentation approaches. Zhang et al. [12] used multiscale Hessian matrix analysis and thresholding for choroidal vessel detection and segmentation, while Mazzaferri et al. [13] and Hu et al. [14] used a graph search approach for choroid layer segmentation. Recently, deep-learning based approaches have gained great interest in medical image segmentation and have demonstrated better performance than traditional image-processing approaches. U-Net [15] is an approach that was developed for the segmentation of biomedical images. Although it has been well demonstrated on 2D medical data [16], [17], the conventional U-Net approach cannot be directly used on volumetric 3D data. A typical strategy is to segment each cross-sectional slice independently of the other slices, and then combine the segmentations to generate a 3D outcome. However, as the segmentations are performed independently, the amalgamated volumetric result can appear disjointed as inter-slice information and continuities are not considered. 3D U-Net [18] addresses this issue by performing a series of 3D convolutional operations on sub-blocks of a volumetric image. Other methods such as V-Net [19] and 3D Deep Supervision Network (3D DSN) [20] also addressed this issue in a similar way using 3D convolution networks. However, these methods require extensive computational memory. While this can be addressed by using Recurrent Neural Networks (RNNs) based approaches to consider volumetric medical images as sequential frames for segmentation, this is computationally intensive and prone to memory leakages.

Motivated by the above observations, we introduce a multi-task learning approach for segmentation of the choroid. Inter-slice spatial information is extracted as a separate task, which is used to reconstruct a slice and segment a target slice. Experiments on a large dataset of volumetric OCT images from a high myopia cohort show that our proposed approach achieves better choroid segmentation with reduced computation complexity. The rest of this paper is organized as follows: Section 2 reviews related works; Section 3 describes the proposed approach; Section 4 describes the experimental results, and Section 5 concludes the paper.

## 2. Related works on segmentation

Medical image segmentation plays an essential role in computer-aided diagnosis systems in different clinical applications [21–24] and many automated segmentation approaches have been developed to delineate region-of-interests for clinical evaluation and diagnosis. The most commonly used convolutional neural network (CNN) architecture for segmentation of two-dimensional(2D) medical images is U-Net [15]. Due to its ability to segment images efficiently with limited amount of training data, the U-Net architecture has been successfully demonstrated in different fields including brain and liver 2D image segmentation. Volumetric segmentation can be performed using U-Net by applying the approach on cross-sectional slices individually, from which the segmentations are amalgamated. For example, CorneaNet [17] is an architecture based on 2D U-Net to perform segmentation of the cornea. Çiçek et al. [18] proposed a 3D

Biomedical Optics EXPRESS

variant of U-Net. A similar approach was also proposed by Milletari et al. [19] in developing V-Net. The main difference between3D U-Net and V-Net is that V-Net incorporated local residual architecture in addition to skip connections in each block. 3D Deep Supervision Network (3D DSN) was developed by Dou et al. [20] for volume-to-volume segmentation by combining 3D convolutional networks with a deep supervision mechanism. However, these approaches usually require the whole volumetric image to be considered, which is computationally expensive with extensive memory requirements. This memory issue can be solved by passing sub-volumes to the architecture instead of the whole volume at the same time. However, this results in discontinuities when we perform volumetric reconstruction from the sub-blocks.

Recently, recurrent networks have been gaining popularity as a sequential approach for volumetric medical image segmentation. Chen et al. [25] and Tseng et al. [26] proposed a combination of FCN (Fully Convolutional Networks) and RNN (Recurrent Neural Networks). However, the inputs to these networks were still based on the full volumetric image. To address this issue, a hybrid combination of sequential processing with 2D segmentation in medical images was proposed by Poudel et al. [27] and Novikov et al. (Sensor3D) [28]. These methods are computationally intensive and are also prone to memory leakages, which are common in recurrent networks. Cahyo at al. [29] addressed the problem of memory leakage by changing the time distributed layers in Sensor3D using a 3D convolution approach. However, this approach was still computationally expensive.

D-UNet, which was proposed by Zhou et al. [30], learnt spatial context from adjacent slices during the encoding stage using 3D convolution, which was combined with 2D segmentation that considered the adjacent slices as different channels. Fang et al. [31] also proposed learning spatial context in the same manner using a Globally Guided Progressive Fusion Network (GGPF-Net). GGPF-Net locally learnt spatial context from patched neighbouring slices, which was then combined with the globally learnt features from the central slice in the bottleneck. These methods outperformed sequential and 3DFCN approaches in terms of performance and computation efficiency.

Other approaches have included an end-to-end multi-task learning that incorporated an attention module to learn specific-task features proposed by Liu et al. [32] and a multi-task learning architecture to reconstruct the foreground and background of the segmentation labels proposed by Chen et al. [33]. Other architectures have also integrated Generative Adversarial Networks (GAN) into the segmentation architecture, such as proposed by Xu et al. [34] and Zhao et al. [35].

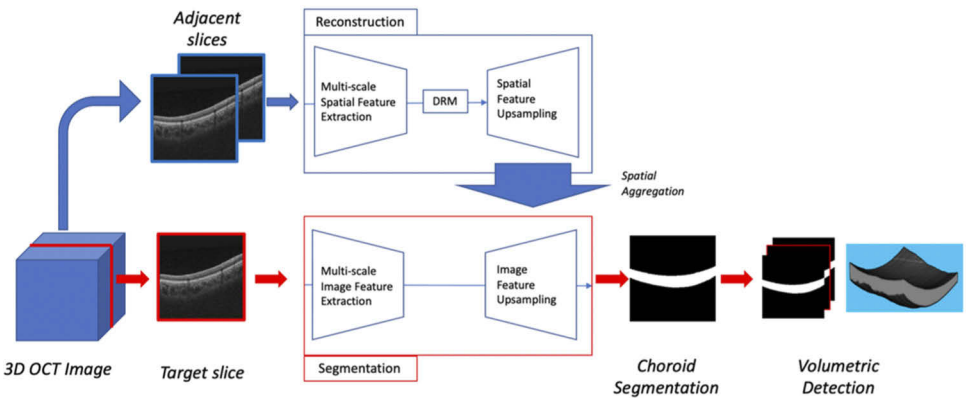


**Fig. 1.** Schematic diagram of the proposed SA-Net. SA-Net learns spatial information from two or more adjacent slices and infuse the spatial information to segmentation task.

Biomedical Optics EXPRESS

Compared to these previous approaches, we propose a novel multi-task learning architecture that learns the spatial information between adjacent slices to reconstruct a selected central slice. We use hard parameter sharing to channel the spatial information between the reconstruction and segmentation tasks. This mechanism makes the model aggregate the spatial features more explicitly since it directly learns the correlation between adjacent slices and the slice that will be segmented. We name this architecture Spatial Aggregated Networks (SA-Net) due to its aggregation of spatial information. Figure 1 shows a schematic view of SA-Net. Spatial context-infused segmentation of each cross-sectional slice is first performed, after which the segmentations are used to construct a volumetric representation of the choroid.

## 3. Methods

The proposed SA-Net architecture incorporates both reconstruction and segmentation tasks. Hard parameter sharing, which is a commonly used technique in multi-task learning architecture to share feature extractor layers, is used because the two tasks require two different inputs, which are co-dependent. This allows the reconstruction task to learn to extract useful spatial features which are shared with the segmentation task. As illustrated in Fig. 1 the segmentation branch performs the 2D segmentation and concurrently, the spatial information extracted from the adjacent slices in the reconstruction branch is infused with the segmentation branch by element-wise addition. Specifically, given a slice, $\boldsymbol{I_i}$, to be segmented, the reconstruction branch will take $\wedge = \{\boldsymbol{I_{i-n}}, \boldsymbol{I_{i+n}}\}$ for its input whereby $\boldsymbol{n}$ defines the appropriate distance of adjacent slices to be used. In this paper we used $\boldsymbol{n} = 5$ after hyperparameter tuning. For both reconstruction and segmentation branch, we used batch normalization for regularization in each block as shown in Fig. 2 which can speed up the training and avoid convergence issues [36].

### 3.1. Reconstruction branch

As shown in Fig. 2, the reconstruction branch can be divided into down-sampling and up-sampling parts. In the down-sampling part, we exploit the rich spatial information from adjacent slices by using 3D convolution and max pooling layers to reduce the features tensor size. This part of the branch enables the architecture to learn local spatial information contained between slices, which are the inter-slice features. To reduce the number of parameters required by 3D convolution layers, we incorporate dimension reduction mechanism (DRM) in the bottleneck block to convert 3D information into 2D information. The converted 3D information is up-sampled using up-sampling layers and 2D convolution layers in the up-sampling part. DRM is also used in the skip connections in the up-sampling part, then the dimensionally squeezed features tensors are concatenated to their corresponding up-sampling layers.

DRM starts with a 3D convolution layer, followed by summing the numerical features along the cross-sectional axis, batch normalization and Rectified Linear Unit (ReLU) activation. This mechanism ensures that the features of the volumetric nature of adjacent slices are retained and at the same time scaled or normalized. Meanwhile, the dimensionality of the tensor is squeezed and the complexity of the reconstruction branch is reduced with increasing converging speed.

After up-sampling, a final 2D convolution is performed and the loss between output and the groundtruth (the $I_i$ slice) is calculated. We used mean squared error to calculate the similarity distance between the predicted output ($y^{pred}$) and groundtruth ($y^{true}$).

$$L_{reconst} = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left[y_{i,j}^{pred} - y_{i,j}^{true}\right]^2 \tag{1}$$

### 3.2. Segmentation branch

The segmentation branch takes the corresponding slice to be segmented as the input. A series of 2D convolution operations followed by max-pooling are performed in the down-sampling

Biomedical Optics EXPRESS

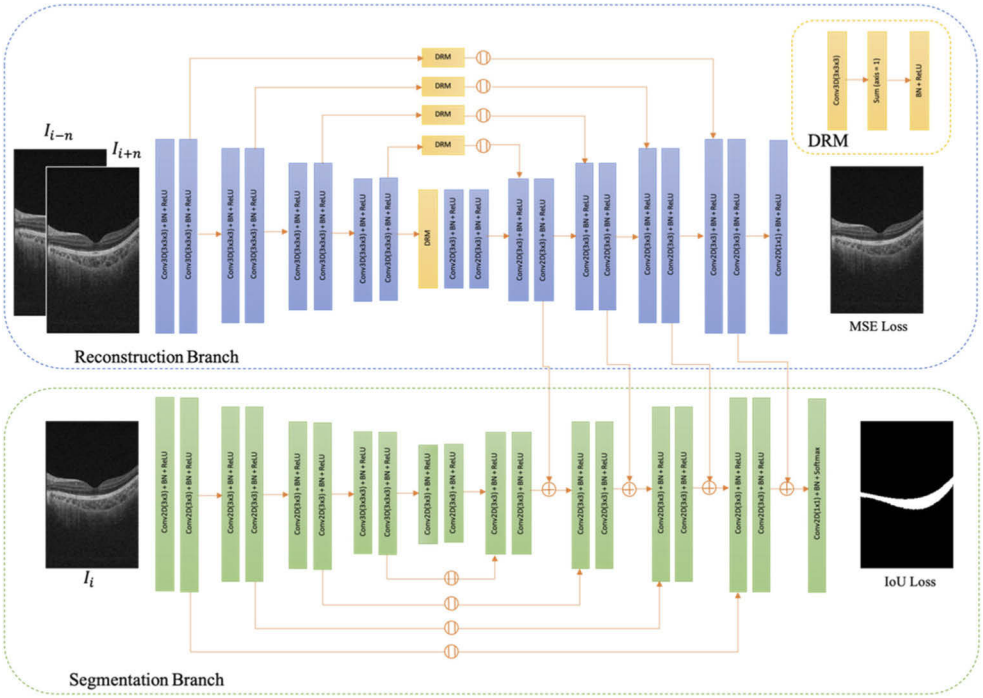


**Fig. 2.** The proposed SA-Net consists of reconstruction and segmentation branches for two different tasks to learn inter-slice and intra-slice features respectively. Spatial information aggregated in reconstruction branch is fused with the segmentation branch during the encoding stage.

part to extract intra-slice features contained within the slice and reduce the features tensor size. In the up-sampling part, we concatenate high-resolution features during down-sampling with low-resolution features.

Each up-sampling block consists of one 2D up-sampling layer and two 2D convolution layers. At each end of these blocks, we fused the knowledge of the inter-slice features from the reconstruction branch. The high-resolution 2D volumetric features are summed with the 2D intra-slice extracted features using element-wise addition to incorporate the inter-correlation features between slices.

The up-sampling branch is ended with a $1 \times 1$ 2D convolution and a sigmoid activation function. We used 2D Intersection over Union (IoU) loss function to maximize the intersection region between the predicted probability map and the groundtruth, as defined in Eq. (2)

$$L_{IoU} = 1 - \frac{\sum_{i,j \in \mathbb{N}} y_{i,j}^{pred} y_{i,j}^{true}}{\sum_{i,j \in \mathbb{N}} y_{i,j}^{pred} + y_{i,j}^{true} - y_{i,j}^{pred} y_{i,j}^{true}} \tag{2}$$

Given the maximized probability map of the segmentation, $y^{pred}$, we threshold this predicted probability map to get the final segmentation result.

$$y_{i,j}^{seg} = \begin{cases} 0, & 1 - y_{i,j}^{pred} \geq \delta \\ 1, & 1 - y_{i,j}^{pred} < \delta \end{cases} \tag{3}$$

where $y^{seg}$, $y^{pred}$ and $\delta$ are the segmentation result, probability map and threshold value respectively. In this paper we used a threshold value 0.5.

Biomedical Optics EXPRESS

### 3.3. Training

Hard parameter sharing is used in this architecture. Instead of constraining layers with regularization in the loss function (soft parameter sharing), during backpropagation, parameters are shared between the two branches during training. The loss function used is the combination of Eq. (1) and Eq. (2).

$$\boldsymbol{L}(\boldsymbol{y}^{\boldsymbol{pred}} \cup \boldsymbol{y}^{\boldsymbol{rec}} | \boldsymbol{W}) = \boldsymbol{L}(\boldsymbol{y}^{\boldsymbol{rec}} | \boldsymbol{W}^{\boldsymbol{rec}}) + \boldsymbol{L}(\boldsymbol{y}^{\boldsymbol{pred}} | \boldsymbol{W}^{\boldsymbol{rec}} \cup \boldsymbol{W}^{\boldsymbol{seg}}) \tag{4}$$

This loss function allows us to update the weights in the reconstruction branch that is also useful for the segmentation task while keeping the weights updated in the segmentation branch to only focus on the segmentation task. This is important to keep the multi-task learning architecture focused on segmentation.

### 3.4. Data acquisition

In this paper we evaluated SA-Net on a high myopia dataset. The high myopia data set was composed of 166 high myopia eyes acquired using a commercial swept-source OCT (SS-OCT) system from 99 patients with refractive error of $-5.18 \pm 2.21$ D, DRI OCT Triton (Topcon Corp., Japan) with a 1050 nm wavelength, scanning speed of 100,000 A-scans/sec and 7 mm $\times$ 7 mm scanning protocol, centered at the macula. Each volumetric image contains 256 cross-sectional with dimensions $256 \times 128$ pixels. The SingHealth Centralized Institutional Review Board approved all protocols, and all methods adhered to the tenets of the Declaration of Helsinki.

We generated the groundtruth for Triton datasets by manual annotation graded by one trained grader. The choroid was determined by evaluating the boundary between the choroid and the RPE layer as well as the choroid tissue with the sclera.

After resizing and normalization, the customized data generator produced two inputs: a target slice and its adjacent slices. The network received the target slice for segmentation together with the adjacent slices as inputs for reconstruction. Slices from the ends of the volume are padded by averaging the target slice with the available adjacent slices.

### 3.5. Experimental design

We used a five-fold cross-validation strategy to train and evaluate the proposed model. Stratified sampling was done over the choroidal volume for each fold to ensure that a similar distribution was achieved and to avoid dataset bias. To further avoid training bias and risk of overfitting, we ensured that all images from the same eye were in the same fold. This avoids a scenario where the testing and training partitions could potentially consist of different images from the same eye. The overall experimental result was then obtained by averaging over all validation sets in each fold.

The results obtained from our proposed SA-Net architecture was compared with five architectures, namely 2D U-Net [15], 3D U-Net [18], BC U-Net [29], Sensor3D [28]. These architectures were trained for 20 epochs with early stopping that will stop the training if the validation loss is not decreasing anymore. Hyperparameters tuning was done for all five architectures, to ensure best performance for individual architectures. Specifically for SA-Net we used batch size equals 5, with image dimension $256 \times 128$, filter size of $3 \times 3$ and individual number of filters of [16,32, 64, 128, 256] (increasing in the downsampling blocks and decreasing in the upsampling blocks). We trained these architectures using Adam [37] with a fixed initial learning rate of 0.001.

The architecture was implemented using Python version3.7.4 and TensorFlow [38] version 2.0 on a workstation with GPU NVIDIA RTX 2080Ti and 64GB RAM.

Biomedical Optics EXPRESS

### 3.6. Evaluation metrics

We evaluated the segmentation result volumetrically by calculating the IoU and Dice score with respect to the groundtruth segmentation.

$$IoU = \frac{\sum_{i,j,k\in\mathbb{N}} y^{seg}_{i,j,k} y^{true}_{i,j,k}}{\sum_{i,j,k\in\mathbb{N}} y^{seg}_{i,j,k} + y^{true}_{i,j,k} - y^{seg}_{i,j,k} y^{true}_{i,j,k}} \tag{5}$$

$$Dice = \frac{2 \times \sum_{i,j,k\in\mathbb{N}} y^{seg}_{i,j,k} y^{true}_{i,j,k}}{\sum_{i,j,k\in\mathbb{N}} y^{seg}_{i,j,k} + y^{true}_{i,j,k}} \tag{6}$$

These metrics measure the performance of the segmentation result over the groundtruth and thus, the ability to differentiate the object from the background, where $y^{seg}$ is the segmentation result as defined in Eq. (3).

We also assessed the inter-slice segmentation by using the choroidal thickness map generated from the choroidal segmentation. The choroidal thickness map was obtained by summing the enface thickness of the detected choroidal layer as shown in Fig. 3. We evaluated the generated map by calculating the SSIM, which assesses the similarity of the predicted thickness map and groundtruth thickness map. Given two images with the same dimension, $x$ and $y$, SSIM formula is given by Eq. (7). [39]

$$SSIM = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{7}$$

where $\mu_x$, $\mu_y$, $\sigma_x$, $\sigma_y$, $\sigma_{xy}$ are the average of $x$, the average of $y$, the variance of $x$, the variance of $y$, and the covariance of x and y respectively. While $c_1 = (0.01DR)^2$ and $c_2 = (0.03DR)^2$ with $DR$ or dynamic range is defined by:

$$DR = \max(y^{true}_{thickness\ map}) - \min(y^{true}_{thickness\ map}) \tag{8}$$

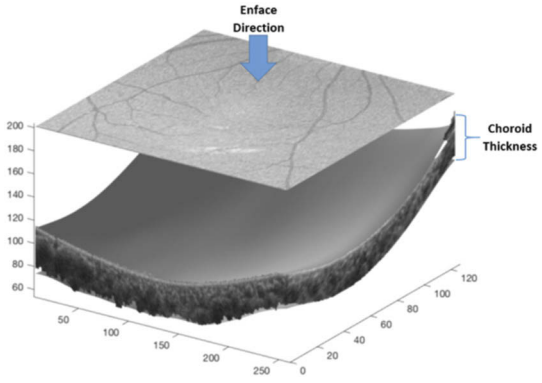


**Fig. 3.** Reconstructed volumetric representation of choroid segmentation. A choroidal thickness map can be generated by measuring the choroidal thickness at each A-scan across the image.

To evaluate the performance comparison between architectures, paired t-test were used for each individual volume. In this paper, $p$-value below 0.05 is considered to be significant.

Biomedical Optics EXPRESS

## 4. Experiment results

### 4.1. Results

Table 1 shows the result of volumetric choroidal segmentation. Results of the segmentations from other approaches are also presented in Table 1 for comparison. Except for SSIM, the average IoU, and the Dice score were measured volumetrically using Eq. (5) and Eq. (6). The volumetric IoU and Dice scores indicate that SA-Net achieved significantly better segmentations ($p$<0.001) compared to other approaches.

**Table 1. Results for Volumetric Choroidal Segmentation on the Triton High Myopia Dataset.[a]**

| Method | IoU | Dice | SSIM | Time/Vol(s) |
|---|---|---|---|---|
| SA-Net | 0.9469 ± 0.0261 | 0.9725 ± 0.0142 | 0.7211 ± 0.0672 | 1.0812 |
| 2D U-Net | 0.9342 ± 0.0306$^{\#}$ | 0.9657 ± 0.0168$^{\#}$ | 0.6924 ± 0.0658$^{\#}$ | 0.2882 |
| 3D U-Net | 0.9251 ± 0.0321$^{\#}$ | 0.9608 ± 0.0179$^{\#}$ | 0.6679 ± 0.0759$^{\#}$ | 0.6533 |
| BC U-Net | 0.9416 ± 0.0299$^{\#}$ | 0.9697 ± 0.0164$^{\#}$ | 0.7108 ± 0.0637$^{\#}$ | 3.1000 |
| Sensor3D | 0.9431 ± 0.0285$^{\#}$ | 0.9705 ± 0.0156$^{\#}$ | 0.6975 ± 0.0671$^{\#}$ | 1.9390 |

[a]Paired t-tests were used to evaluate if differences were significant;
* and $^{\#}$ indicate $p$-value < 0.05 and $p$-value < 0.001.

It can also be observed that the SSIM performance for SA-Net significantly ($p$<0.001) outperformed BC U-Net, which shows that both approaches were able to generate thickness maps which were close to that of the groundtruth. However, SA-Net also required less time for training and inference compared to BC U-Net due to the reduced network complexity and computational requirements. The upper and lower boundary error was measured and also shown in Table 2. We can see that SA-Net outperformed all of the other architectures, although only by small margin with respect to Sensor3D architecture.

**Table 2. Upper and Lower Boundary Error for Volumetric Choroidal Segmentation on the Triton High Myopia Dataset.[a]**

| Method | Mean Absolute Upper Boundary Error(mm) | Mean Absolute Lower Boundary Error(mm) |
|---|---|---|
| SA-Net | 0.0022 ± 0.0014 | 0.0103 ± 0.1172 |
| 2D U-Net | 0.0037 ± 0.0019$^{\#}$ | 0.0121 ± 0.0141$^{\#}$ |
| 3D-U-Net | 0.0047 ± 0.0019$^{\#}$ | 0.0124 ± 0.0078$^{\#}$ |
| BC U-Net | 0.0025 ± 0.0015$^{\#}$ | 0.0109 ± 0.0094* |
| Sensor3D | 0.0023 ± 0.0016 | 0.0106 ± 0.0061 |

[a]Paired t-tests were used to evaluate if differences were significant;
* and $^{\#}$ indicate $p$-value < 0.05 and $p$-value < 0.001.

Figure 4 shows the examples of the 2D cross-sectional segmentation result for each architecture, where the red line indicates the segmentation result and the green line is the groundtruth. It can be observed that cross-sectional segmentation of the choroid using SA-Net is better than that from the other architectures and generates a smoother segmentation result.

Figure 5 shows the choroidal thickness map generated from the groundtruth and various segmentation approaches. We observed that the thickness map from the 2D U-Net is more noisy with higher variability between slices, while the patch-based 3D U-Net resulted in a more patchy result with discontinuities (indicated by the black arrows) between patches (sub-volumes). BC U-Net, Sensor3D, and SA-Net showed results, which were more similar to the groundtruth. SA-Net was shown to perform significantly better than other approaches for the SSIM score as shown in Table 1, and by visual comparison with the groundtruth choroidal thickness map.

Biomedical Optics EXPRESS

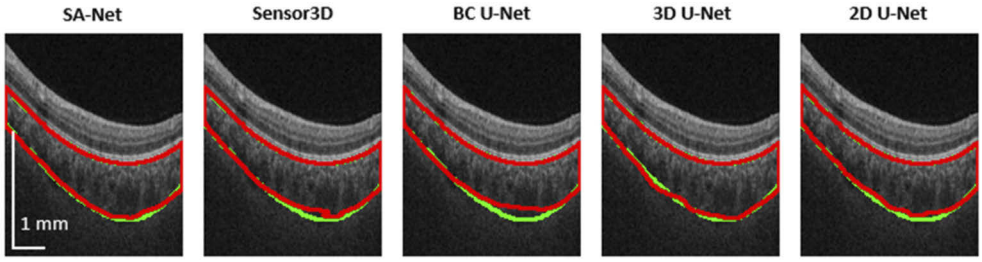


**Fig. 4.** Comparison of choroidal segmentation in cross-sectional images using different approaches. Green and red lines indicate groundtruth and segmentation, respectively. The choroidal segmentation using SA-Net can be observed to be closer to the other approaches.

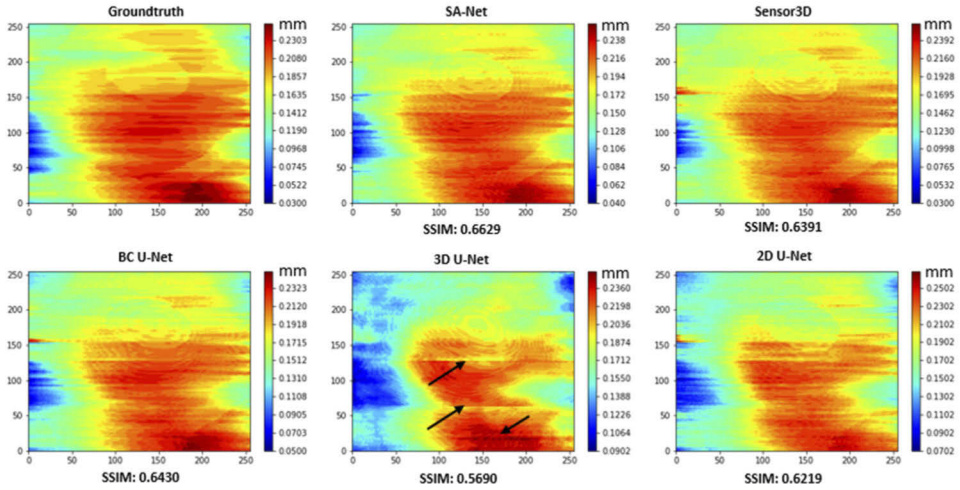


**Fig. 5.** Choroid layer thickness map for each architectures. SA-Net shows the most similar result with the groundtruth. The quantitative similarity is calculated using SSIM. X and Y axes indicate image dimension in pixel, while the colorbars indicate the thickness in millimeters.

Although the results show that the proposed SA-Net outperforms other segmentation approaches, a limitation of the results is that an independent test set was unavailable for testing. However, the same-folds were used for all methods, and hyperparameter tuning was performed to optimize each of the architectures.

### *4.2. Evaluation on various choroid thickness level*

In eyes with more severe myopia, the thinned choroids can present a challenge and may be more difficult to segment than thicker choroids. To assess if choroidal thickness affected the performance of SA-Net, we divided the Triton dataset based on their sub-foveal choroidal thickness, which is defined as the choroidal thickness below fovea. Tan et al. [40] defined eyes with a sub-foveal thickness of less than 300 $\mu m$ as thin choroids, while intermediate choroids were those with choroidal thickness between $301\mu m - 400\mu m$ inclusive. Using these criteria, we considered the analysis separately for 126 thin choroids and 40 intermediate choroids.

Figure 6 depicts an example of the segmentation results for images with thin and intermediate choroids. Generally, segmentation of the thin choroids is more difficult than intermediate choroids

Biomedical Optics EXPRESS

due to the thinner choroidal profiles and higher variation in the choroidal visibility. SA-Net showed good performance across both choroid thickness types.

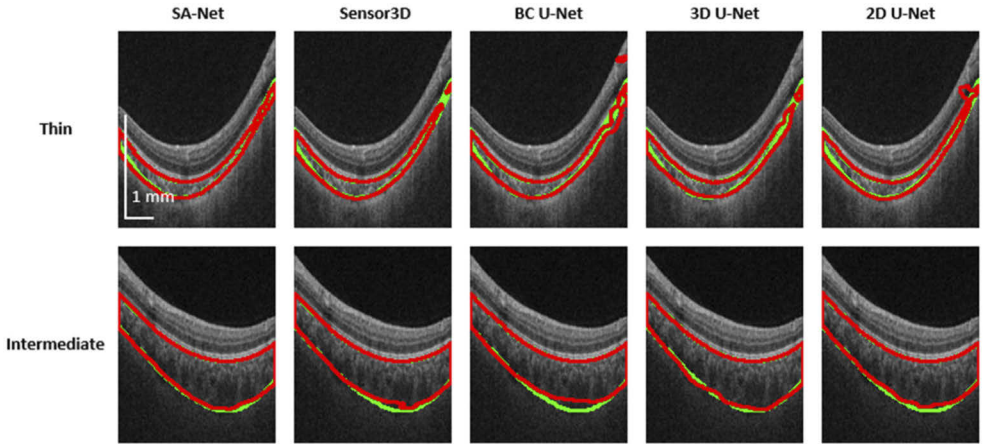


**Fig. 6.** Example choroidal segmentation results on cross-sectional images from eyes with thin and intermediate choroid thickness. Green and red lines indicate groundtruth and segmentations respectively. The results show that SA-Net achieved a better segmentation of thin choroids. In intermediate choroids, the segmentations were more similar.

In Table 3 and Table 4 we can see the performance comparisons thin and intermediate choroids separately. SA-Net performed significantly better ($p<0.001$) compared to the other architectures for thin choroids. For intermediate choroids, although SA-Net performed significantly better ($p<0.05$) than the other approaches, the differences in the segmentation performance was smaller compared to the thin choroids.

**Table 3. Results for Volumetric Choroidal Segmentation for Thin Choroids.[a]**

| Method | IoU | Dice | SSIM |
|---|---|---|---|
| SA-Net | $0.9431 \pm 0.0315$ | $0.9705 \pm 0.0171$ | $0.7133 \pm 0.0739$ |
| 2D U-Net | $0.9288 \pm 0.0370^{\#}$ | $0.9628 \pm 0.0205^{\#}$ | $0.6870 \pm 0.0738^{\#}$ |
| 3D U-Net | $0.9188 \pm 0.0387^{\#}$ | $0.9574 \pm 0.0216^{\#}$ | $0.6530 \pm 0.0860^{\#}$ |
| BC U-Net | $0.9379 \pm 0.0356^{\#}$ | $0.9677 \pm 0.0196^{\#}$ | $0.7039 \pm 0.0704^{\#}$ |
| Sensor3D | $0.9391 \pm 0.0335^{\#}$ | $0.9683 \pm 0.0183^{\#}$ | $0.6911 \pm 0.0729^{\#}$ |

[a]Paired t-tests were used to evaluate if differences were significant;
[a]$*$ and $^{\#}$ indicate $p$-value $< 0.05$ and $p$-value $< 0.001$.

### 4.3. *Discussion on the complexity of the network and memory usage*

Table 1 shows Time/Vol which is a measure of the inference time for each architecture. This is a measure of the network complexity as a more complicated network will require a longer inference time. However, architectures such as 3D U-Net need to be assessed not only based on their complexity but also on their memory usage during training and inference.

To measure the memory usage of each architecture, we calculated the number of the parameters computed during training or inference. Table 5 shows the number of the features and the trainable parameters for each architecture. Features refer to the output of the filters, while trainable parameters refer to the architecture's parameters. Complexity of the network is determined by

Biomedical Optics EXPRESS

**Table 4. Results for Volumetric Choroidal Segmentation for Intermediate Choroids.[a]**

| Method | IoU | Dice | SSIM |
|---|---|---|---|
| SA-Net | $0.9589 \pm 0.0347$ | $0.9790 \pm 0.0186$ | $0.7455 \pm 0.1440$ |
| 2D U-Net | $0.9514 \pm 0.0176^{\#}$ | $0.9750 \pm 0.0094^{\#}$ | $0.7093 \pm 0.0798^{\#}$ |
| 3D U-Net | $0.9448 \pm 0.0281^{\#}$ | $0.9716 \pm 0.0151^{\#}$ | $0.7136 \pm 0.1203^{\#}$ |
| BC U-Net | $0.9554 \pm 0.0229^{\#}$ | $0.9770 \pm 0.0126^{\#}$ | $0.7450 \pm 0.0669$* |
| Sensor3D | $0.9558 \pm 0.0453$* | $0.9773 \pm 0.0248$* | $0.7175 \pm 0.1522^{\#}$ |

[a]Paired t-tests were used to evaluate if differences were significant;
[a]* and # indicate $p$-value $< 0.05$ and $p$-value $< 0.001$.

trainable parameters. Networks such as 2D U-Net and 3D U-Net have low complexity, while BC U-Net and Sensor3D have highly complex architectures.

**Table 5. Number of Element of the Features and the Trainable Parameters.**

| Method | Features ($\times 10^6$) | Parameters ($\times 10^6$) | Total Parameters ($\times 10^6$) |
|---|---|---|---|
| SA-Net | 46.94 | 5.5 | 52.44 |
| 2D U-Net | 18.46 | 1.87 | 20.43 |
| 3D U-Net | 772.34 | 5.89 | 778.23 |
| BC U-Net | 57.85 | 10.35 | 68.20 |
| Sensor3D | 90.52 | 10.36 | 100.88 |

3D U-Net has a very large number of features compared to other architectures. This results in large memory requirements, which is a major limitation in the use of 3D U-Net. In contrast, although 2D U-Net has much fewer features, our results show that it does not perform as well in volumetric segmentation as compared to the other architectures, that incorporate spatial information. Compared to the other architectures, SA-Net has a faster inference time and fewer parameters.

## 5. Conclusion

Spatial information can provide useful context for volumetric segmentation. Our proposed architecture, SA-Net incorporates spatial information from corresponding adjacent slices to explicitly integrate spatial correspondences. We compared SA-Net with other recent approaches for segmenting the choroid in volumetric OCT images on a high myopia dataset, and demonstrated that SA-Net outperformed the other approaches in segmentation accuracy and quality of the generated choroidal thickness map, with lesser complexity and memory usage. The analysis across various choroid thicknesses also showed that SA-Net performed particularly well for challenging thin choroids. Our results show that SA-Net could be used for efficient and accurate segmentation of OCT data, and can be useful for monitoring choroidal changes in highly myopic eyes. SA-Net could also be potentially applied to other types of volumetric medical images.

**Funding.** National Medical Research Council (CG/C010A/2017_SERI, MOH-000249-00, MOH-OFIRG20nov-0014, OFIRG/0048/2017, OFLCG/004c/2018, TA/MOH-000249-00/2018); Singapore Eye Research Institute & Nanyang Technological University (SERI-NTU Advanced Ocular Engineering Program); National Research Foundation Singapore (NRF2019-THE002-0006, NRF-CRP24-2020-0001); Agency for Science, Technology and Research (A20H4b0141); SERI-Lee Foundation (LF1019-1); Duke-NUS Medical School (Duke-NUS-KP(Coll)/2018/0009A).

**Disclosures.** The authors declare no conflicts of interest.

**Data availability.** Data underlying the results presented in this paper are not publicly available at this time due to privacy reasons.

Biomedical Optics EXPRESS

## References


1. C. M. Yancey and R. A. Linsenmeier, "Oxygen distribution and consumption in the cat retina at increased intraocular pressure," Invest. Ophthalmol. Visual Sci. **30**(4), 600–611 (1989).
2. J. Chhablani, I. Y. Wong, and I. Kozak, "Choroidal imaging: a review," Saudi J. Ophthalmol. **28**(2), 123–128 (2014).
3. S. A. Read, J. A. Fuss, S. J. Vincent, M. J. Collins, and D. A. Caneiro, "Choroidal changes in human myopia: insights from optical coherence tomography imaging," Clin. Exp. Optometry **102**(3), 270–285 (2019).
4. T. Fujiwara, Y. Imamura, R. Margolis, J. S. Slakter, and R. F. Spaide, "Enhanced depth imaging optical coherence tomography of choroid in highly myopic eyes," Am. J. Ophthalmol. **148**(3), 445–450 (2009).
5. I. Flores-Moreno, F. Lugo, J. S. Duker, and J. M. Ruiz-Moreno, "The relationship between axial length and choroidal thickness in eyes with high myopia," Am. J. Ophthalmol. **155**(2), 314–319.e1 (2013).
6. R. F. Spaide, "Age-related choroidal atrophy," Am. J. Ophthalmol. **147**(5), 801–810 (2009)..
7. J. Y. Lee, D. H. Lee, J. Y. Lee, and Y. H. Yoon, "Correlation between subfoveal choroidal thickness and the severity or progression of nonexudative age-related macular degeneration," Invest. Ophthalmol. Visual Sci. **54**(12), 7812–7818 (2013)..
8. J. T. Kim, D. H. Lee, S. G. Joe, J. G. Kim, and Y. H. Yoon, "Changes in choroidal thickness in relation to the severity of retinopathy and macular edema in type 2 diabetic patients," Invest. Ophthalmol. Visual Sci. **54**(5), 3378–3384 (2013).. PMID: 23611988.
9. K. Devarajan, R. Sim, J. Chua, C. W. Gong, S. Matsumura, H. M. Htoon, L. Schmetterer, S. M. Saw, and M. Ang, "Optical coherence tomography angiography for the assessment of choroidal vasculature in high myopia," Br. J. Ophthalmol. **104**(7), 917–923 (2020).
10. D. P. Popescu, L. Choo-Smith, C. Flueraru, Y. Mao, S. Chang, J. Disano, S. Sherif, and M. G. Sowa, "Optical coherence tomography: fundamental principles, instrumental designs and biomedical applications," Biophys. Rev. **3**(3), 155–169 (2011)..
11. M. Ang, C. W. Wong, Q. V. Hoang, G. C. M. Cheung, S. Y. Lee, A. Chia, S. M. Saw, K. Ohno-Matsui, and L. Schmetterer, "Imaging in myopia: potential biomarkers, cuurent challenges and future developments," Br. J. Ophthalmol. **103**(6), 855–862 (2019).
12. L. Zhang, K. Lee, M. Niemeijer, R. F. Mullins, M. Sonka, and M. D. Abràmoff, "Automated segmentation of the choroid from clinical SD-OCT," Invest. Ophthalmol. Visual Sci. **53**(12), 7510–7519 (2012)..
13. J. Mazzaferri, L. Beaton, G. Hounye, D. N. Sayah, and S. Constantino, "Open-source algorithm for automatic choroid segmentation of OCT volume reconstructions," Sci. Rep. **7**(1), 42112 (2017)..
14. Z. Hu, X. Wu, Y. Ouyang, Y. Ouyang, and S. R. Sadda, "Semiautomated segmentation of the choroid in spectral-domain optical coherence tomography volume scans," Invest. Ophthalmol. Visual Sci. **54**(3), 1722–1729 (2013)..
15. O. Ronnerberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer, 2015), pp. 234–241.
16. J. Kugelman, D. Alonso-Caneiro, S. A. Read, J. Hamwood, S. J. Vincent, F. K. Chen, and M. J. Collins, "Automatic choroidal segmentation in OCT images using supervised deep learning methods," Sci. Rep. **9**(1), 13298 (2019)..
17. V. Aranha, L. Schmetterer, H. Stegmann, M. Pfister, A. Messner, G. Schmidinger, G. Garhofer, and R. M. Werkmeister, "CorneaNet: fast segmentation of cornea OCT scans of healthy and keratoconic eyes using deep learning," Biomed. Opt. Express **10**(2), 622–641 (2019).
18. Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: learning dense volumetric segmentation from sparse annotation," in *International Conference on Medical Image Computing and Computer Assisted Intervention* (Springer, 2016), pp. 424–432.
19. F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: fully convolutional neural networks for volumetric medical image segmentation," in *International Conference on 3D Vision* (2016) pp. 565–571,.
20. Q. Dou, H. Chen, Y. Jin, L. Yu, J. Qin, and P.-A. Heng, "3D deeply supervised network for automatic liver segmentation from CT volumes," in *International Conference on Medical Image Computing and Computer Assisted Intervention* (Springer, 2016), pp. 149–157.
21. L. Dora, S. Agrawal, R. Panda, and A. Abraham, "State-of-the-art methods for brain tissue segmentation: a review," IEEE Rev. Biomed. Eng. **10**, 235–249 (2017)..
22. L. Lenchik, L. Heacock, A. A. Weaver, R. D. Boutin, T. S. Cook, J. Itri, C. G. Filippi, R. P. Gullapalli, J. Lee, M. Zagurovskaya, and T. Retson, "Automated segmentation of tissues using CT and MRI: a systematic review," Academic Radiol. **26**(12), 1695–1706 (2019).
23. A. P. Yow, R. Srivastava, J. Cheng, A. Li, J. Liu, L. Schmetterer, H.L. Tey, and D. W. K. Wong, "Techniques and applications in skin OCT analysis," Adv. Exp. Med. Biol. **1213**, 149–163 (2020).. PMID: 32030669.
24. M. D. Abràmoff, M. K. Garvin, and M. Sonka, "Retinal imaging and image analysis," IEEE Rev. Biomed. Eng. **3**, 169–208 (2010).. PMID: 22275207; PMCID:PMC3131209.
25. J. Chen, L. Yang, Y. Zhang, M. Alber, and D. Z. Chen, "Combining fully convolutional and recurrent neural networks for 3D biomedical image segmentation," *Advances in Neural Information Processing Systems* (2016) pp. 3036–3044.
26. K.-L. Tseng, Y.-L. Lin, W. Hsu, and C.-Y. Huang, "Joint sequence learning and cross-modality convolution for 3D biomedical segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition* (2017).

Biomedical Optics EXPRESS


27. R. P. K. Poudel, P. Lamata, G. Montana, K. Bhatia, B. Kainz, M.H. Moghari, and D. F. Pace, "Recurrent fully convolutional neural networks for multi-slice MRI cardiac segmentation," *Reconstruction, Segmentation, and Analysis of Medical Images*, Springer, pp. 83–94, 2017.
28. A. A. Novikov, D. Major, M. Wimmer, D. Lenis, and K. Bühler, "Deep sequential segmentation of organs in volumetric medical scans," IEEE Trans. Med. Imaging **38**(5), 1207–1215 (2019).
29. D. A. Y. Cahyo, D. W. K. Wong, A. P. Yow, S.-M. Saw, and L. Schmetterer,"Volumetric choroidal segmentation using sequential deep learning approach in high myopia subjects," in *IEEE/Engineering in Medicine and Biology Conference*, Canada, 2020.
30. Y. Zhou, W. Huang, P. Dong, Y. Xia, and S. Wang, "D-UNet: a dimension-fusion U shape network for chronic stroke lesion segmentation," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, doi: 10.1109/TCBB.2019.2939522.
31. C. Fang, G. Li, C. Pan, Y. Li, and Y. Yu, "Globally guided progessive fusion network for 3D pancreas segmentation," in *International Conference on Medical Image Computing and Computer Assisted Intervention* (Springer, 2019), pp. 210–218.
32. S. Liu, E. Johns, and A. J. Davison, "End-to-end multi-task learning with attention," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 1871–1880, 2019.
33. S. Chen, G. Bortsova, A. G.-U. Juárez, G. van Tulder, and M. de Bruijne, "Multi-task attention-based semi-supervised learning for medical image segmentation," in *International Conference on Medical Image Computing and Computer Assisted Intervention* (Springer, 2019), pp. 457–465.
34. K. Xu, L. Wen, G. Li, L. Bo, and Q. Huang, "Spatiotemporal CNN for video object segmentation," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. USA, pp. 1379–1388, 2019.
35. M. Zhao, L. Wang, J. Chen, D. Nie, Y. Cong, S. Ahmad, A. Ho, P. Yuan, S. H. Fung, H. H. Deng, J. Xia, and D. Shen, "Chraniomaxillo-facial bony structures segmentation from MRI with deep-supervision adversarial learning," in *International Conference on Medical Image Computing and Computer Assisted Intervention* (Springer, 2018), pp. 720–727.
36. S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *proceedings of the32nd International Conference on International Conference on Machine Learning*, Volume 37 (ICML'15) (2015), 448–456.
37. D. Kingma and J. Ba, "Adam: a method for stochastic optimization," in ICLR, 2015.
38. M. Abadi, P. Barham, E. Brevdo, Z. Chen, C. Citro, G.S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Schlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Watterberg, M. Wicke, Y. Wu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015. Software available from tensorflow.org.
39. Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," J. Lightwave Technol. **29**(4), 439–448 (2011).
40. C. S. Tan, Y. Ouyang, H. Ruiz, and S. R. Sadda, "Diurnal variation of choroidal thickness in normal, healthy subjects measured by spectral domain optical coherence tomography," Invest. Ophthalmol. Visual Sci. **53**(1), 261–266 (2012)..